

Teaching Commentary on “A Primary Care-Based Multidisciplinary Readmission Prevention Program”: Essential Aspects of Comparability and Context in Practice-Based Program Evaluation

Elizabeth M. Yano, PhD, MSPH^{1,2}

¹VA HSR&D Center for the Study of Healthcare Innovation, Implementation and Policy, VA Greater Los Angeles Healthcare System, Sepulveda, CA, USA; ²Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, USA.

J Gen Intern Med 29(5):805–7

DOI: 10.1007/s11606-014-2820-2

© Society of General Internal Medicine 2014

INTRODUCTION

In their paper evaluating a primary care-based multidisciplinary readmission prevention program, Cavanaugh and colleagues describe the quality improvement processes they underwent to develop, iteratively refine, implement and then evaluate a program designed to improve post-discharge planning and care transitions, thereby reducing readmissions.¹ Using a retrospective cohort study design, they then describe the results of their program evaluation comparing patients who have been discharged from their hospital and exposed to a local innovation—a hospitalization follow-up clinic—to a local comparison group matched on readmission risk and admission timing. In unadjusted analyses, program participants had significantly fewer readmissions at 30 and 90 days, with non-significant trends for lower emergency department visits, shorter median time to first post-discharge follow-up visit, and higher achievement of hospital follow-up within 30 days.¹ When they adjusted for baseline differences between their intervention and comparison groups, only the hazard of a 90-day readmission was significantly lower in the intervention group.

This paper provides an excellent example of the potential for and challenges of identifying an appropriate and comparable comparison group in non-experimental study designs, as well as the value of what Jubelt and colleagues described as “thick descriptions” of program content and processes of implementation, in context such that actions and behaviors become clear to outsiders.² In this commentary, I lay out the implications of both comparability and context for evaluating quality improvement (QI) programs, with an eye to their contribution to implementation and spread of local innovations.

Comparability of Comparisons: Enhancing Internal Validity

While even randomization does not always confer comparability, the utility and internal validity of non-experimental evaluation designs often rise or fall on the basis of the comparability of the comparison group(s) used.³ The purpose of a comparison group is to offer evidence of what would have happened in the absence of the intervention to the extent possible, enabling estimation of an incremental advantage attributable to program implementation.⁴ Therefore, the type and manner in which selection criteria are applied, the availability of data and measures, and the covariate adjustments that evaluators deploy to increase the comparability of intervention vs. comparison groups become essential design elements.⁵ Otherwise, potential threats to internal validity begin to mount, undermining confidence in the study’s findings.⁶

Cavanaugh and colleagues¹ compared two groups of patients that had been discharged from their hospital: (1) the intervention group, members of whom had been exposed to a hospitalization follow-up clinic; and (2) a comparison group, derived using the same inclusion/exclusion criteria, matched on a general readmission risk stratification score, and whose index discharge was within a month of the intervention patient to whom they were matched. To their credit, Cavanaugh et al.’s approach reduced many common validity threats by applying comparable criteria and similar time trajectories (e.g., within a month of index discharges) for both groups over the same 5-month time period and in the same facility.

At the same time, each of these strengths may be moderated by one or more challenges (Table 1). For example, use of the same facility means that patients in both intervention and comparison groups shared the same area and organizational environments, which reduces the need to adjust for contextual factors. However, implementing a program in the same site simultaneously increases the potential for contamination, as patients, providers and staff with experience in the intervention

Table 1. Strengths and Challenges Associated with Comparison Group Methods in Cavanaugh et al.¹

Selected Attributes	Strength	Challenge
Comparison group from same facility	<ul style="list-style-type: none"> • Groups share same environment • ↓ need to adjust for contextual factors 	<ul style="list-style-type: none"> • ↑ potential for contamination • ↓ external validity given single facility reliance
Use of concurrent comparison group	<ul style="list-style-type: none"> • ↓ influence of seasonality in staff availability and access • ↓ history threats 	<ul style="list-style-type: none"> • Selection threats in unblinded setting (unconscious influences on program referrals⁶)
Same inclusion/exclusion criteria applied	<ul style="list-style-type: none"> • ↑ comparability as members have been subjected to the same eligibility requirements 	<ul style="list-style-type: none"> • ↓ representation of population at risk for readmission • ↓ utility and external validity for patients deemed ineligible
Assignment to comparison group from eligible pool	<ul style="list-style-type: none"> • Random sample of patients without exposure to program 	<ul style="list-style-type: none"> • Inherent selection effects in program referral, attendance, participation, completion
Matched on risk strata related to outcome of interest	<ul style="list-style-type: none"> • ↑ comparability of inherent risk of readmission between groups 	<ul style="list-style-type: none"> • ↓ comparability due to unmeasured risks • Matching precluded risk adjustment in models
Comparison group member's index discharge within a month of matched intervention patient	<ul style="list-style-type: none"> • ↑ comparability of time to event or outcome of interest 	<ul style="list-style-type: none"> • Time differential in matching may still explain differences in time to follow-up

may co-mingle or otherwise influence referrals, participation or behavior. Single-site studies also limit external validity, but all studies have to start somewhere, and a rush to multi-site research without solid pilot evidence would be unwarranted, not to mention prohibitively expensive. The authors also applied the same inclusion/exclusion criteria, increasing the comparability of intervention and comparison patients. Yet only patients with established primary care providers were eligible, among other exclusions, limiting knowledge about how the program might be deployed or tailored for other types of patients. And while comparison group patients were randomly sampled from the pool of eligible patients, the factors driving program referral (let alone attendance/participation) produce inherent selection effects.⁷ Matching comparison group patients to individual intervention patients based on the risk of readmission using a locally standardized risk classification scheme also represented a methodological strength. The evaluators' approach to risk stratification and matching yielded intervention and comparison group members of low, moderate and high risk in equal proportions, seeking comparability in a fundamental and inherent driver of readmissions. At the same time, despite best efforts, the intervention group ultimately included significantly more patients with chronic obstructive pulmonary disease

(COPD) or asthma (46 % vs. 24 %) and fewer with cirrhosis (4 % vs. 22 %) or depression (22 % vs. 44 %).¹ As is commonly done, the authors adjusted for these and other baseline differences in their final analyses, as they focused on better understanding the program's independent effects.⁷ The extent to which the baseline differences continued to be associated with program participation after statistical adjustment is not reported; understanding whether these differences had meaningful implications for their findings may have been a useful adjunct, especially for patients with mental health conditions.

That said, the reality is that for every strength, any number of challenges is always present. Criteria that seek to homogenize patients into comparable, highly-controlled groups to improve internal validity often limit our ability to translate research and QI evidence into routine clinical practice.^{8,9} Therefore, achieving comparability of intervention and comparison groups in program evaluation requires constant balancing of the demands of both internal and external validity.

Implications of Context for Considering External Validity, Implementation and Spread

A central tenet of external validity is the ability to generalize beyond the circumstances of a particular study to other patients, providers and practices.¹⁰ However, local innovations, like the case presented here, are just that—local—and reflect the environment within which providers practice and patients experience care, and within which quality improvement and evidence-based programs may or may not take hold and work.¹¹ As a result, area and organizational contexts have substantial influences on quality of care, the design and deployment of QI interventions, and, ultimately, the implementation of evidence-based practice and policy.¹²

Cavanaugh and colleagues¹ again provide a strong example for describing the local context within which they developed, tested and evaluated their follow-up program for individuals at risk for hospital readmission. They describe their organizational context as a large, academic internal medicine practice based at the University of North Carolina (UNC) at Chapel Hill. The practice is further distinguished by a 15-year history of QI activities, reflected by apparent in-house QI capabilities (e.g., use of root cause analysis, process mapping, small tests of change, run charts) and their referenced peer-reviewed publications, as well as NCQA recognition as a Level 3 PCMH. This latter recognition reflects the highest level a practice can achieve on standards of access/continuity, population management, planning and care management, self-management support/community resources, tracking/coordinating care, and performance measurement/improvement.¹³ Not surprising for a

Level 3 PCMH practice, they also enjoy the benefits of an electronic health record (EHR), which clearly enabled access to data for both comparison and intervention patients. They also provide caseload and workload parameters: they serve approximately 14,000 patients, about 270 of whom are admitted to UNC Hospitals on a monthly basis and about 20 % (54) of whom, in turn, are readmitted within 30 days on average.

This description provides an extremely useful backdrop for understanding context and its implications for external validity, which in turn informs how readily other practices might be able to adopt (or need to adapt) the hospitalization follow-up clinic as a model. First, to what other types of practices might the evaluation results generalize? A major advantage is its reliance on practice-based research. As Lawrence Green notes, "If we want more evidence-based practice, we need more practice-based evidence."¹⁴ While this study was not part of a practice-based research network (PBRN), the practice under study appears to have operated under similar precepts: use of the practice as a "laboratory" for testing healthcare innovations with engagement of frontline providers and staff in the framing and design of potential solutions to problems with which they have experience, benefiting from elements of QI and participatory action research.^{15,16} As an academic practice, however, it may be difficult to generalize to non-academic settings, where the majority of Americans obtain their health care. The multi-level partnerships between the UNC practice and UNC Hospitals may also not be easily reproducible elsewhere, yet are essential to implementation and spread of evidence-based practice.^{17,18} EHR implementation is also very uneven in the U.S., with significant organizational determinants of variation in use and quality of decision support and data mining.

Adoption decisions by other organizations may be predicated on the extent to which being an enriched, academic practice actually served as a core attribute of the intervention and its implementation. While not defined as such in the paper, in "realistic evaluations," interventions are conceptualized and evaluated as being inextricably linked to their contexts, such that "mechanisms [of change] may be more or less effective in producing their intended outcomes, depending on their interaction with various contextual factors."^{19,20} For this innovation to have impacts beyond its relatively narrow context, readers may therefore need to consider how they might have to adapt their practice and/or the innovation to fit their own organizational milieu.^{21,22}

Corresponding Author: Elizabeth M. Yano, PhD, MSPH; VA HSR&D Center for the Study of Healthcare Innovation, Implementation and Policy, VA Greater Los Angeles Healthcare System, 16111 Plummer Street, Mailcode 152, Sepulveda, CA 91343, USA (e-mail: Elizabeth.yano@va.gov).

REFERENCES

1. Cavanaugh JJ, Jones CD, Embree G, Tsai K, Miller T, Shilliday BB, McGuirt B, Roche R, Pignone M, DeWalt DA, Ratner MD. A primary care-based multidisciplinary readmission prevention program. *J Gen Intern Med*. 2014 (in press).
2. Jubelt LE, Metlay JP, Kravitz RL. Implementation science workshop: an open call for cases for a new JGIM feature bridging research and practice. *J Gen Intern Med*. 2013;28(11):1383-4.
3. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston, MA: Houghton Mifflin Publishers; 1979.
4. McDavid JC, Hawthorn LRL. Program evaluation and performance measurement. Thousand Oaks, CA: Sage Publications; 2006.
5. Crombie IK, Davies HTO. Research in health care: design, conduct and interpretation of health services research. Wiley & Sons: New York, NY; 1996.
6. Grembowski D. The practice of health program evaluation. Thousand Oaks, CA: Sage Publications; 2001.
7. DesJarlais DC, Lyles C, CRepez N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94(3):361-6.
8. Rubenstein LV, Pugh J. Strategies for promoting organizational and practice change by advancing implementation research. *J Gen Intern Med*. 2006;21(Suppl 2):S58-64.
9. Westfall JM, Mold J, Fagnan L. Practice-based research—"Blue Highways" on the NIH roadmap. *JAMA*. 2007;297(4):403-6.
10. Green LW, Glasgow R. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval Health Prof*. 2006;29(1):126-53.
11. Chaudoir SR, Dugan AG, Barr CH. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. *Implement Sci*. 2013;8:22.
12. Yano EM. Influence of health care organizational factors on implementation research: QUERI Series. *Implement Sci*. 2008;3:29.
13. National Committee for Quality Assurance. Patient-centered medical homes. www.ncqa.org/programs/recognition/patientcenteredmedicalhomePCMH.aspx. Accessed November 25, 2013.
14. Green LW. Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *Am J Public Health*. 2006;96(3):406-9.
15. Green LA, Hickner J. A short history of primary care practice-based research networks: from concept to essential research laboratories. *J Am Board Fam Med*. 2006;19(1):1-10.
16. Peters DH, Adam T, Alonge O, Agypong A. Implementation research: what it is and how to do it. *BMJ*. 2013;347:1-7.
17. Yano EM, Green LW, Glanz K, Ayanian J, Mittman BS, Chollette V, Rubenstein LV. Implementation and spread of multi-level interventions in practice: implications for the cancer care continuum. *JNCI*. 2012;44:86-99.
18. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet*. 2003;362(9391):1225-30.
19. Rycroft-Malone J, Fontenla M, Bick D, Seers K. A realistic evaluation: the case of protocol-based care. *Implement Sci*. 2010;5:38.
20. Best A, Greenhalgh T, Lewis S, Saul JE, Carroll S, Bitz J. Large-system transformation in health care: a realist review. *Milbank Q*. 2012;90(3):421-56.
21. Kochevar LK, Yano EM. Understanding health care organization needs and context: beyond performance gaps. *J Gen Intern Med*. 2006;21:S25-9.
22. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013;13(6S):S23-30.

Conflict of Interest: The author declares that she has no conflict of interest.